OPEN ACCESS

KING'S College LONDON

King's Research Portal

DOI:
10.2337/db16-0315

Document Version
Peer reviewed version

Link to publication record in King's Research Portal

Citation for published version (APA):
Arredouani, A., Stocchero, M., Culeddu, N., El-Sayed Moustafa, J., Tichet, J., Balkau, B., Brousseau, T., Manca, M., Falchi, M., & D.E.S.I.R. study group (2016). Metabolomic Profile of Low Copy-Number Carriers at the Salivary Alpha-Amylase Gene Suggests a Metabolic Shift Towards Lipid-Based Energy Production. *Diabetes*. https://doi.org/10.2337/db16-0315

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

•Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
•You may not further distribute the material or use it for any profit-making activity or commercial gain
•You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

METABOLOMIC PROFILE OF LOW COPY-NUMBER CARRIERS AT THE SALIVARY ALPHA-AMYLASE GENE SUGGESTS A METABOLIC SHIFT TOWARDS LIPID-BASED ENERGY PRODUCTION

Abdelilah Arredouani[1,4,¢,✉] , Matteo Stocchero[2,¢], Nicola Culeddu[3], Julia El-Sayed Moustafa[4], , the D.E.S.I.R. study group, Jean Tichet[11], Beverley Balkau[6,7], Thierry Brousseau[5], Marco Manca[8,9], Mario Falchi[4,10,✉]

[1] Hamad Ben Khalifa University, Qatar Biomedical Research Institute, Diabetes Research Centre, Qatar Foundation, Doha, Qatar

[2] S-IN Soluzioni Informatiche, Vicenza, Italy

[3] Istituto di Chimica Biomolecolare, Lab. NMR, CNR, v. La Crucca 3, 07040 Sassari, Italy

[4] Department of Twin Research & Genetic Epidemiology, King's College London, St. Thomas' Hospital Campus, London, UK

[5]UF8832 - Biochimie Automatisée, Pôle de Biologie Pathologie Génétique, Centre Hospitalier Régional Universitaire, Lille, France

[6] Centre de Recherche en Epidémiologie et Santé des Populations, CESP, INSERM U1018, Renal and cardiovascular epidemiology, Villejuif, France

[7] Universities Versailles-St Quentin and Paris-Sud, UMRS 1018, Villejuif, France

[8] Experimental Vascular Pathology, CARIM, University of Maastricht, NL

[9] ATS-DO, CERN, Geneva, CH

[10] Department of Genomics of Common Disease, Imperial College London, UK

[11] IRSA, La Riche, France

¢ These authors contributed equally

✉ Corresponding authors

Diabetes Publish Ahead of Print, published online July 19, 2016

Abstract

Low serum salivary amylase levels have been associated with a range of metabolic abnormalities, including obesity and insulin resistance. We recently suggested that low copy-number at the *AMY1* gene, associated with lower enzyme levels, also increases susceptibility to obesity. To advance our understanding of the effect of *AMY1* copy-number variation on metabolism, we compared the metabolomic signatures of high and low copy-number carriers. We analysed, using mass spectrometry and NMR, the sera of healthy normal-weight women carrying either low (LA:≤4 copies; *n*=50) or high (HA:≥8 copies; *n*=50) *AMY1* copies. Best fitting multivariate models (empirical $P<1x10^{-3}$) of MS and NMR data were concordant in showing differences in lipid metabolism between the two groups. In particular, LA carriers showed lower levels of long- and medium-chain fatty acids, and higher levels of dicarboxylic fatty acids and 2-hydroxybutyrate (known marker of glucose malabsorption). Taken together, these observations suggest increased metabolic reliance on fatty acids in LA carriers through β- and ω-oxidation and reduced cellular glucose uptake with consequent diversion of acetyl-CoA into ketogenesis. Our observations are in line with previously-reported delayed glucose uptake in LA carriers after starch consumption. Further functional studies are needed to extrapolate from our findings to implications for biochemical pathways.

Salivary α-amylase (sAA), encoded by the *AMY1* gene, is produced under the control of the autonomic nervous system by the salivary glands. It initiates starch digestion in the mouth, yielding a mixture of maltose, isomaltose and glucose. Therefore, it was expected that individuals with high levels of sAA would show higher postprandial glycaemia, compared to low sAA counterparts. A recent study showed, surprisingly, the opposite effect (1). Lower serum amylase levels have been associated with susceptibility to metabolic abnormalities, including non-alcoholic fatty liver disease, obesity, and diabetes (2-4).

Salivary amylase level is strongly correlated with the copy-number (CN) at the *AMY1* gene (5). Higher *AMY1* copy-number and sAA protein levels were observed in populations traditionally consuming starch-rich diets compared to those consuming a low-starch diet (5). We recently confirmed *AMY1* gene expression and protein levels in serum to be inversely correlated with obesity and suggested CN variation at *AMY1* to be an important driver of this relationship, showing association between low amylase CN and increased BMI and obesity risk in both adults (6) and children (7). On the other hand, a novel study has suggested that CNV might not be the main factor driving the association between amylase levels and obesity susceptibility (8). The mechanisms by which sAA contributes to an increased risk of metabolic diseases remain elusive (8,9). To further investigate the role of genetic variation at the *AMY1* gene on the metabolism, we used untargeted ^{1}H-Nuclear Magnetic Resonance (NMR) and Mass Spectrometry (MS)-based metabolomics, and compared the metabolomic profiles of a carefully selected sample of 100 healthy and normal-weight, age-matched women carrying either low or high *AMY1* CN.

RESEARCH DESIGN AND METHODS

Study population and design.

Only women were included to provide a more homogenous study population to maximize power, given our small study population. The characteristics of the women carrying high (HA;n=50) or low (LA;n=50) copies of the *AMY1* gene are shown in Table S1. The women were selected from the prospective D.E.S.I.R. cohort (9) from among those for whom data on the inferred CN at the *AMY1* gene were already available (6). To

control for the effects of potential confounders, we carefully selected two subsamples from the top and bottom 20% of the CN distribution (≤4; ≥8 CN). Exclusion criteria included: 1) BMI>25kg/m^2 or <18.5kg/m^2; 2) fasting glycaemia>6 mmol/l; 3) development of any metabolic disorders during 9-year follow-up period. Food and caloric intake were not significantly different between the HA and LA groups.

All participants signed a written informed consent and the study was approved by the Ethics Committee for the Protection of Subjects for Biomedical Research of Bicêtre Hospital (France).

Serum amylase levels

Serum pancreatic and total amylase levels were measured by enzymatic colorimetric assays with an auto-Roche/Hitachi cobas c701 (AMY-P-20766623322 and AMYL2-03183742122, Hoffman–La Roche). The sAA levels were obtained by subtracting the pancreatic amylase level from the total amylase level.

Mass spectrometry

MS-based metabolomics profiling was carried out at Metabolon (Durham, NC, USA). The processing and analysis of samples was performed as previously described (10) (Supplementary data).

^{1}H-Nuclear Magnetic Resonance

Spectra of serum samples were measured at 600.13 MHz on a Bruker Avance II 600-MHz spectrometer (Bruker BioSpin GmbH Rheinstetten, Germany) following Beckoner et al (11). NMR spectra were phased and baseline corrected. Spectra were aligned using the CluPA algorithm (12) and the alanine signal used as reference (1.4784 ppm). Residue water signal was removed. Data were reduced to 513 bins by intelligent bucketing, and normalized by Total Sum Normalization. Mean centering and Pareto scaling were applied prior to performing the data analysis.

Identification of metabolomics signatures

We employed multivariate statistical data analyses based on projection methods. Principal Component Analysis (PCA) was used for the initial exploratory data analysis. Metabolomics signatures distinguishing

HA and LA women were identified through Projection to Latent Structures Discriminant Analysis, based on VIP (Variable Importance in Projection) selection (PLS-DA VIP-based) (13). Given the limited sample size, we used a careful and conservative approach to validate our results.

First, results were validated through N-fold full cross-validation, using different values of N (N=6,7,8; we report only results for 7-fold as Q^2 and AUC_{CV} - two measures to estimate the predictive ability of the model), as well as through 1,000 permutation testing on the HA/LA classes (13).

Second, to avoid overfitting we performed stability selection by Monte-Carlo subsampling: real differences should be present consistently, and therefore should be found even under perturbation of the data by subsampling (14). We generated 500 random datasets by Monte Carlo subsampling (with prior probability of 0.70), and then PLS-DA VIP-based was applied to each subsample, obtaining a set of 500 discriminant models. Each model was additionally validated by N-fold full cross-validation and 1,000 permutations of the class response. Metabolites discriminating between HA and LA were identified as those that were selected by >50% of all the models.

Both in the original and in the randomly generated datasets, the VIP threshold for the selection of metabolites discriminating between HA and LA classes was determined by maximising Q^2, while the number of latent variables of the PLS model was calculated by maximising Q^2 under the constraint to pass the permutation test. To clarify model interpretation, PLS-DA VIP-based models were post-transformed into the equivalent Orthogonal-PLS-DA models (OPLS-DA) where the predictive and the orthogonal part of the models were identified.

RESULTS

Serum amylase levels

HA subjects showed significantly higher levels of serum sAA compared to LA counterparts (Wilcoxon $P=6.04\times10^{-8}$). A smaller difference was detected for the pancreatic amylase levels (Wilcoxon $P=0.025$), which is likely to be a reflection of both physiological and structural correlations between the two enzymes (15) (Table S1;Figure S1).

Model Fitting and Validation

The analysis of the NMR and MS datasets by PCA detected the presence of 15 outliers that were excluded from further analysis. The major patterns within the original data were captured by a small number of components, three for both NMR and MS datasets. The score scatter plots in Figure 1A-1B show the separation of the HA and LA samples based on the PLS-DA analyses of the MS and NMR data. The goodness of fit and the predictive ability estimated by 7-fold cross-validation were R^2=0.45 and AUC_{CV}=0.79 for the NMR dataset and R^2=0.71 and AUC_{CV}=0.92 for the MS dataset. Figures 1C-1D compare the R^2 and Q^2 obtained through random permutation of the class response with what was observed in the original dataset (empirical P<0.001 for both R^2 and Q^2).

Stability selection and identification of the HA/LA metabolomics signature

All datasets generated through Monte Carlo sampling successfully passed both permutation testing and N-fold full cross-validation, thus suggesting that model fitting in the data was not driven by a small number of observations. On the basis of the stability selection, we identified two signatures from MS and NMR, encompassing 41 metabolites and 15 signals, respectively (Tables 1-2) that provided the largest contribution to the discrimination of the HA and LA groups. To investigate the role of each metabolite in the discrimination of the two groups, we plotted the median of the VIP calculated by stability selection *versus* the correlation between the predictive latent variable of the OPLS-DA model and the measured variables. Since the VIP and the correlation coefficient estimate the strength and the direction of the effect respectively, metabolites in the right quadrant of the plot showed higher levels in the HA group, while those in the left quadrant higher levels in the LA group (Figures 2A-2B).

Metabolites from MS data were mapped into the Ingenuity Pathway Analysis Knowledge Base database (QIAGEN, Redwood City, CA, USA; accessed October 2015). The reference set included the endogenous chemicals; we chose to focus only on experimentally observed interactions with no restrictions on cell type or species. The top 5 most significantly enriched Molecular and Cellular Functions included Carbohydrate metabolism, Energy production, and Lipid metabolism (Table S2). Their most significant functional annotations were: Oxidation of glucose-6-phosphate (P=3.12×10^{-8}), Oxidation of lipid

(P=1.32×10^{-6}), Accumulation of acylglycerol (P=6.19×10^{-8}), and triacylglycerol (P=2.85×10^{-6}). Of the 41 MS compounds, 28 (68%) were fatty acids (FAs), 17 and 11 with respectively higher and lower levels in LA sera, as compared to HA. Similarly, 9 of the 15 NMR signals (60%) corresponded to FAs and their levels were higher in LA sera.

Despite ^{1}H-NMR experiments being highly reproducible, different lipids cannot be easily distinguished from each other, but are rather assigned to multiple macro-categories characterised by the abundance of specific aliphatic chains, and of their magnetic shielding. Therefore, since the signals of the lipids detected by NMR can be characterized by only the three main functional groups (CH2-CH=CH-, -CH2-, and -CH3), we investigated their correlation with the identified MS metabolites. The strongest correlations ($|r|$>0.7) were observed between the four dicarboxylic fatty acids (DFAs) sebacate, azelate, suberate, and dodecanedioate identified by MS and the -CH2- NMR bucket in the region from 1.15 and 1.19 ppm. Interestingly the same DFAs were also strongly correlated (r>0.8) with the glucose NMR bucket at 3.40 ppm. The BMI distribution was not different between the HA and LA groups (Wilcoxon P=0.986). BMI was independent of HA/LA group definition and did not have any effect on the multivariable PLS models (r^2<0.01; P>0.20). We further excluded association between BMI and each single discriminant metabolite through linear regression considering both BMI and HA/LA level as independent variables (P for BMI>0.10).

DISCUSSION

In this study we aimed to explore the effect of copy-number variation at *AMY1* on metabolism by comparing the metabolomic signature in carefully matched healthy women, carrying low or high CN at the *AMY1* gene as reflected in a significant difference in enzyme levels between the two groups (P=6.04×10-8) (fig. S1). While this study design facilitates the identification of discriminant metabolites associated with *AMY1* CN, results might be affected by spectrum bias as women were selected from the two ends of the CN distribution.

The best fitting models (empirical P<1x10^{-3}) identified through multivariate analyses of the MS and NMR metabolomics data concordantly highlighted differences in lipid metabolism between LA and HA

women. Analysis of the MS discriminant metabolites using the Ingenuity Knowledge Base annotations confirmed the significant enrichment of functional categories for “Lipid metabolism”, “Energy production”, and “Carbohydrate metabolism”. In particular, the MS data showed that the levels of four DFAs, suberate, sebacate, azelate, and dodecanedioate, were higher in the sera from LA women compared to HA (Table 2). This increase in DFAs levels is suggestive of an upregulation of the ω-oxidation. In normal conditions, ω-oxidation plays a minor role in overall FAs oxidation. However, in certain physiological states, where mitochondrial β-oxidation is blocked or overwhelmed, free FAs are ω-oxidized in the reticulum, by a CYP-4A dependent pathway, to DFAs (16), which can then enter mitochondrial β-oxidation.

To corroborate the proposed interpretation, alongside the increase in DFAs, women with LA also showed lower serum concentrations of several medium and long chain fatty acids (Table 2). This reduction in FA levels is coherent with a more active uptake and oxidation, both via β- and ω-oxidation.

Interestingly, LA women also showed increased levels of 2-hydroxybutyrate, a biomarker for insulin resistance in humans (17) (Table 2). Taken together, these results suggest a pattern of reduced cellular glucose uptake, and consistent metabolic shift toward lipid exploitation, in LA women. Interestingly, NMR suggested a slight increase of glucose levels in LA sera.

In a similar line, Mandel and colleagues (18) recently published an experiment where they measured the dynamics of postprandial glucose and insulin following controlled starch, or glucose meals in healthy LA and HA carriers. While LA and HA did not show any differences after the glucose-based meal, LA carriers showed significantly higher glucose levels following the starch-based meal, suggesting glucose malabsorption after starch ingestion in people with constitutionally low salivary amylase levels. Mandel et al advocated that this effect might be mediated by delayed preabsorptive insulin response in the LA group after the starch meal.

In summary, our results support the hypothesis that salivary α-amylase CN might play a role in glucose uptake, offering novel evidence to support further research on its mechanisms of action.

ACKNOWLEDGMENTS

AA and MF conceived and designed the study. MS performed the statistical analysis and wrote the respective section. NC performed the NMR data analysis and wrote the respective section. TB performed the measurement of the amylase levels. AA, MF wrote the first draft of the manuscript. AA, MF, MS, MM, NC, JEM, BB helped interpreting the data and contributed to the final version of the manuscript. JT conceived the D.E.S.I.R. protocol, participated in the conduct of the study over 10 years and aided in financing the study. The D.E.S.I.R. Group provided the serum samples and the demographic data. AA and MF are the guarantors of this work and, as such, have full access to the metabolomics data used in this study and take responsibility for the integrity of the data and the accuracy of the data analysis. The authors are grateful to all participants of this study.

No potential conflicts relevant to this article were reported

This study was supported by an intramural grant from the Qatar Biomedical Research Institute, a member of Qatar Foundation.

The D.E.S.I.R. Study Group includes Inserm-U1018 (Paris: B. Balkau, P. Ducimetière, E. Eschwège), Inserm-U367 (Paris: F. Alhenc-Gelas), CHU d'Angers (A. Girault), Bichat Hospital (Paris: F. Fumeron, M. Marre, R. Roussel); CHU de Rennes (F. Bonnet), CNRS UMR-8199 (Lille: S. Cauchi, P. Froguel), Medical Examination Services (Alençon, Angers, Blois, Caen, Chartres, Chateauroux, Cholet, Le Mans, Orléans and Tours), Research Institute for General Medicine (J. Cogneau), General practitioners of the region, and Cross- Regional Institute for Health (C. Born, E. Caces, M. Cailleau, N. Copin, O. Lantieri, J.G. Moreau, F. Rakotozafy, J. Tichet, S. Vol).

The D.E.S.I.R. study has been supported by Inserm contracts with CNAMTS, Lilly, Novartis Pharma and Sanofi-aventis, and by Inserm (Réseaux en Santé Publique, Interactions entre les déterminants de la santé, Cohortes Santé TGIR 2008), the Association Diabète Risque Vasculaire, the Fédération Française de Cardiologie, La Fondation de France, ALFEDIAM, ONIVINS, Société Francophone du Diabète, Ardix Medical, Bayer Diagnostics, Becton Dickinson, Cardionics, Merck Santé, Novo Nordisk, Pierre Fabre, Roche and Topcon.

REFERENCES

1. Villa R, Mandel AL, Jones BD, La Clair JJ, Burkart MD: Structure of FD-895 revealed through total synthesis. Organic letters 2012;14:5396-5399

2. Lee JG, Park SW, Cho BM, Lee S, Kim YJ, Jeong DW, Yi YH, Cho YH: Serum amylase and risk of the metabolic syndrome in Korean adults. Clinica chimica acta; international journal of clinical chemistry 2011;412:1848-1853

3. Nakajima K, Nemoto T, Muneyuki T, Kakei M, Fuchigami H, Munakata H: Low serum amylase in association with metabolic syndrome and diabetes: A community-based study. Cardiovascular diabetology 2011;10:34

4. Nakajima K, Oshida H, Muneyuki T, Saito M, Hori Y, Fuchigami H, Kakei M, Munakata H: Independent association between low serum amylase and non-alcoholic fatty liver disease in asymptomatic adults: a cross-sectional observational study. BMJ open 2013;3

5. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC: Diet and the evolution of human amylase gene copy number variation. Nature genetics 2007;39:1256-1260

6. Falchi M, El-Sayed Moustafa JS, Takousis P, Pesce F, Bonnefond A, Andersson-Assarsson JC, Sudmant PH, Dorajoo R, Al-Shafai MN, Bottolo L, Ozdemir E, So HC, Davies RW, Patrice A, Dent R, Mangino M, Hysi PG, Dechaume A, Huyvaert M, Skinner J, Pigeyre M, Caiazzo R, Raverdy V, Vaillant E, Field S, Balkau B, Marre M, Visvikis-Siest S, Weill J, Poulain-Godefroy O, Jacobson P, Sjostrom L, Hammond CJ, Deloukas P, Sham PC, McPherson R, Lee J, Tai ES, Sladek R, Carlsson LM, Walley A, Eichler EE, Pattou F, Spector TD, Froguel P: Low copy number of the salivary amylase gene predisposes to obesity. Nature genetics 2014;46:492-497

7. Mejia-Benitez MA, Bonnefond A, Yengo L, Huyvaert M, Dechaume A, Peralta-Romero J, Klunder-Klunder M, Garcia Mena J, El-Sayed Moustafa JS, Falchi M, Cruz M, Froguel P: Beneficial effect of a high number of copies of salivary amylase AMY1 gene on obesity risk in Mexican children. Diabetologia 2015;58:290-294

8. Usher CL, Handsaker RE, Esko T, Tuke MA, Weedon MN, Hastie AR, Cao H, Moon JE, Kashin S, Fuchsberger C, Metspalu A, Pato CN, Pato MT, McCarthy MI, Boehnke M, Altshuler DM, Frayling TM, Hirschhorn JN, McCarroll SA: Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. Nature genetics 2015;

9. Balkau B, Eschwege E, Tichet J, Marre M: Proposed criteria for the diagnosis of diabetes: evidence from a French epidemiological study (D.E.S.I.R.). Diabetes & metabolism 1997;23:428-434

10. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E: Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Analytical chemistry 2009;81:6656-6667

11. Beckonert O, Keun HC, Ebbels TM, Bundy J, Holmes E, Lindon JC, Nicholson JK: Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nature protocols 2007;2:2692-2703

12. Vu TN, Valkenborg D, Smets K, Verwaest KA, Dommisse R, Lemiere F, Verschoren A, Goethals B, Laukens K: An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. BMC bioinformatics 2011;12:405

13. Chong I-G, Jun C-H: Performance of some variable selection methods when multicollinearity is present. Chemometrics and Intelligent Laboratory Systems 2005;78:103-112

14. Wehrens R, Franceschi P, Vrhovsek U, Mattivi F: Stability-based biomarker selection. Analytica chimica acta 2011;705:15-23

15. Carpenter D, Dhar S, Mitchell LM, Fu B, Tyson J, Shwan NA, Yang F, Thomas MG, Armour JA: Obesity, starch digestion and amylase: association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes. Human molecular genetics 2015;24:3472-3480

16. Miura Y: The biological significance of omega-oxidation of fatty acids. Proceedings of the Japan Academy Series B, Physical and biological sciences 2013;89:370-382

17. Ferrannini E, Natali A, Camastra S, Nannipieri M, Mari A, Adam KP, Milburn MV, Kastenmuller G, Adamski J, Tuomi T, Lyssenko V, Groop L, Gall WE: Early metabolic markers of the development of dysglycemia and type 2 diabetes and their physiological significance. Diabetes 2013;62:1730-1737

18. Mandel AL, Breslin PA: High endogenous salivary amylase activity is associated with improved glycemic homeostasis following starch ingestion in adults. The Journal of nutrition 2012;142:853-858

Table 1. Metabolites identified by stability selection applied to MS dataset. Upward and downward arrows in the last column indicate that the LA/HA ratio of the metabolite level is increased (↑) or decreased (↓) in sera from LA compared to HA subjects (see also Figure 2).

Metabolitearial	Class	Pathway	LA/HA
N6-acetyllysine	Amino acid	Lysine metabolism	↑
Histidine		Histidine metabolism	↓
2-hydroxybutyrate (AHB)		Cysteine, methionine, SAM, taurine metabolism	↑
Xylitol	Carbohydrate	Nucleotide sugars, pentose metabolism	↑
Ribose			↑
Pyridoxate	Cofactors and vitamins	Pyridoxal metabolism	
			↓
Succinylcarnitine	Energy	Krebs cycle	↓
Sebacate (decanedioate)	Lipid	Fatty acid, dicarboxylate	↑
Azelate (nonanedioate)			↑
Suberate (octanedioate)			↑
Dodecanedioate			↑
3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF)			↓
2-linoleoylglycerol (2-monolinolein)		Monoacylglycerol	↑
1,2-propanediol		Ketone bodies	↑
7-beta-hydroxycholesterol		Sterol/Steroid	↓
Campesterol			↓
Cortisone			↑
Taurolithocholate 3-sulfate		Bile acid metabolism	↓
Oleamide		Fatty acid, amide	↓
10-heptadecenoate (17:1n7)		Long chain fatty acid	↓
Arachidate (20:0)			↓
Myristoleate (14:1n5)			↓
Palmitate (16:0)			↓
Myristate (14:0)			↓
1-linoleoylglycerol (1-monolinolein)		Monoacylglycerol	↑
2-linoleoylglycerophosphocholine		Lysolipid	↑
1-oleoylglycerophosphoethanolamine			↓
2-oleoylglycerophosphoethanolamine			↑
Pelargonate (9:0)		Medium chain fatty acid	↑
Caprate (10:0)			↓
Laurate (12:0)			↓
Glycerol		Glycerolipid metabolism	↓
Linolenate [alpha or gamma; (18:3n3 or 6)]		Essential fatty acid	↓
Stearoyl sphingomyelin		Sphingolipid	↓
13-methylmyristic acid		Fatty acid, branched	↓
N-acetylcarnosine	Peptide	Dipeptide derivative	↓
Leucylglycine		Dipeptide	↓
Leucylphenylalanine			↓
Gamma-glutamylglutamate		gamma-glutamyl	↓
Hwesasxx*		Polypeptide	↑
Inosine	Nucleotide	Purine metabolism	↑

Table 2. Metabolites identified by stability selection applied to NMR dataset. Upward and downward arrows in the last column indicate that the LA/HA ratio of the metabolite level is increased (↑) or decreased (↓) in sera from LA compared to HA subjects (see also Figure 2). The signal of metabolites was defined with reference to alanine (1.4784 ppm).

Metabolite	Class	Pathway	LA/HA	From ppm	To ppm
Urea		Urea cycle	↑	5.25	5.29
D-glucose	Carbohydrate	Glycolysis	↑	3.39	3.39
Lactic acid			↓	1.33	1.34
Lactic acid			↓	1.32	1.33
Glutamate		Glutamate/glutamine metabolism	↓	2.03	2.07
Fatty acids CH2-CH=CH-	Lipid	Lipid metabolism	↑	1.97	2.01
Fatty acids -CH2-			↑	1.28	1.32
Fatty acids -CH2-			↑	1.23	1.27
Fatty acids -CH2-			↑	1.19	1.23
Fatty acids -CH2-			↑	1.15	1.19
Linoleic chain			↑	0.88	0.92
Fatty acids -CH3			↑	0.83	0.88
Fatty acids -CH3			↑	0.75	0.79
Fatty acids -CH3			↑	0.71	0.75

FIGURE LEGENDS

Figure 1. PLS-DA VIP-based models for MS and NMR datasets. The goodness of fit and predictive ability estimated by 7-fold full cross-validation were R^2=0.45, Q^2=0.25 (AUC_{CV}=0.79) for MS dataset and R^2=0.71, Q^2=0.40 (AUC_{CV}=0.92) for NMR dataset. **A** and **B:** score scatter plots obtained after the transformation of the PLS-DA VIP-based models into OPLS-DA models for the MS (**A**) and NMR (**B**) datasets. The samples are coloured according to the class: HA (grey dots) and LA (black dots). The x-axis represents the predictive latent variable tp providing a clear class separation of LA against HA while the y-axis is the first orthogonal latent variable to_1 that is not related to the class samples. **C** and **D:** plot for the response permutation test (1,000 random permutations) of the models built for MS (**C**) and NMR (**D**). The x-axis represents the correlation between the class response Y and the random permutated class response Yp while the y-axis reports the values of R^2 and Q^2 for the calculated PLS-DA VIP-based models. Most of the models having low correlation between permutated response and the class response shows negative Q^2 and the related R^2 is always lower than that calculated for the reference model (i.e. the model corresponding to cor(Yp,Y)=1). This behaviour proves the absence of overfitting in the reported models.

Figure 2. VIP *versus* cor(tp,x) plot for the OPLS-DA models of the MS (**A**) and NMR (**B**) datasets reporting the metabolites identified by stability selection. The x-axis shows the direction of the effects of the metabolites on the model expressed as cor(tp,x) while the y-axis reports the strength of their contribution to the model (VIP): metabolites on the left of the origin are overrepresented in LA (negative values for cor(tp,x)) while metabolites on the right of the origin are overrepresented in HA (positive values for cor(tp,x)); the strength of the metabolite in the separation between LA and HA increases according to the VIP value.

Figure 1

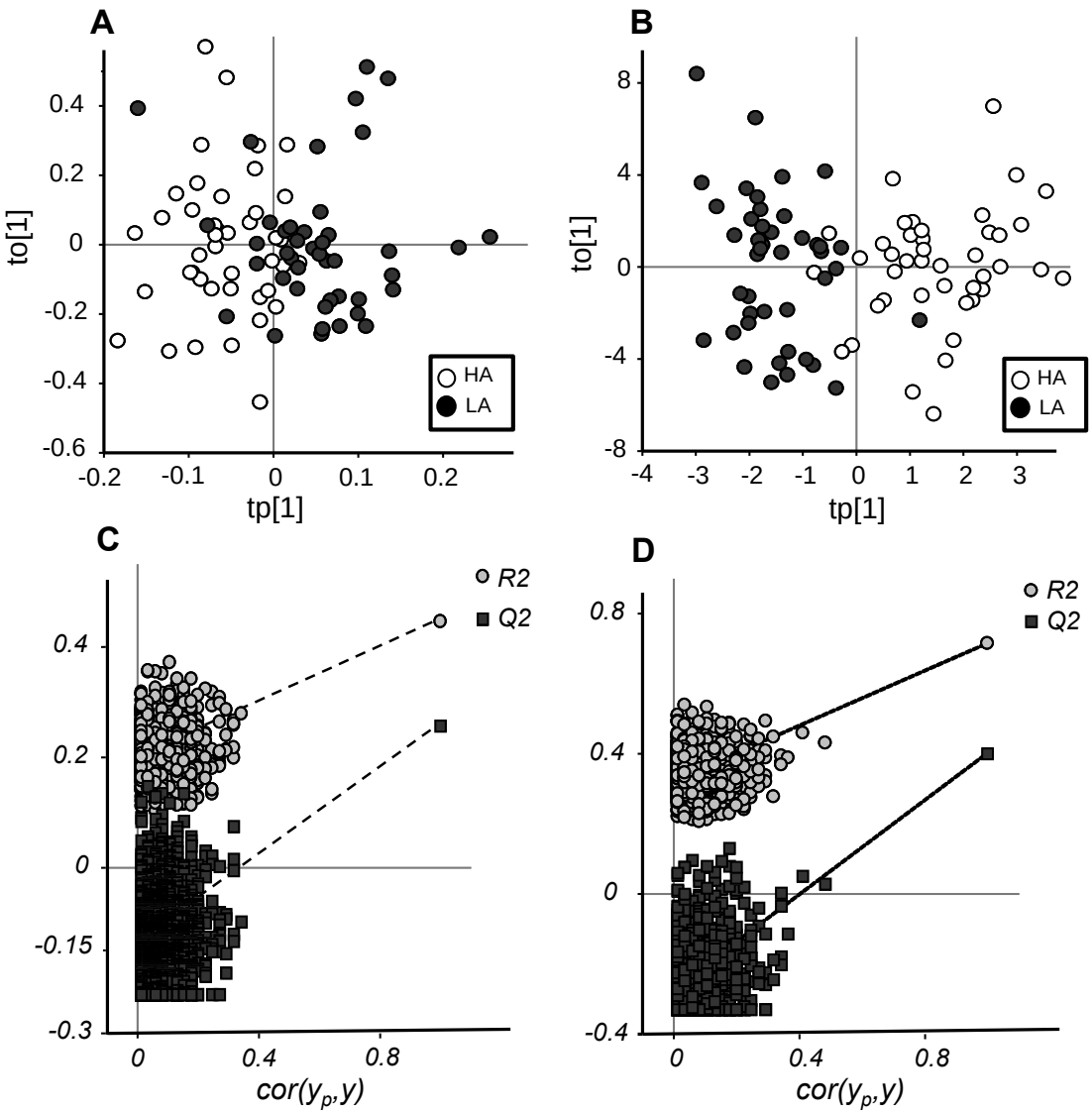
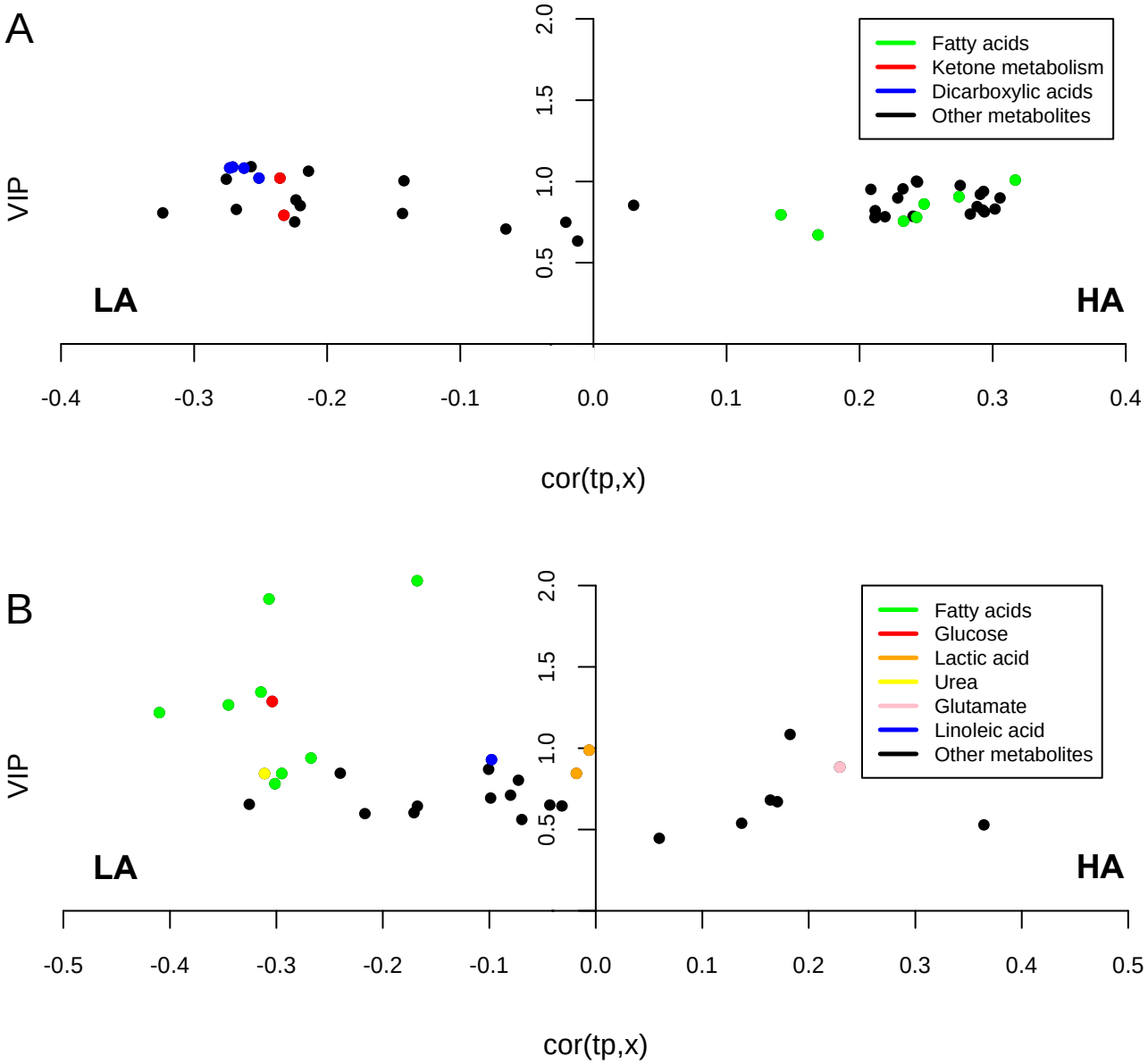


Figure 2


A
VIP
LA
HA
Fatty acids
Ketone metabolism
Dicarboxylic acids
Other metabolites
0.5
1.0
1.5
2.0
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
cor(tp,x)
B
VIP
LA
HA
Fatty acids
Glucose
Lactic acid
Urea
Glutamate
Linoleic acid
Other metabolites
0.5
1.0
1.5
2.0
-0.5
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
cor(tp,x)

SUPPORTING MATERIAL

MATERIALS

SAMPLE SELECTION.

Inferred *AMY1* copy numbers were already available in a sample of 2,137 participants from the D.E.S.I.R. cohort, selected for not showing any evidence of incident diabetes, glucose intolerance, or dyslipidemia over the course of a nine year follow-up (Falchi *et al* 2014). Serum amylase data (total, pancreatic, and salivary) were measured in these subjects.

For this metabolomics analysis, we aimed to select, among those with available *AMY1* CNV data, two homogenous groups who differed only in the number of copies of the salivary amylase gene.

We reasoned that many factors including age, sex, and metabolic disease, may alter metabolite levels as identified through mass spectrophotometry and NMR,, that could therefore act as potential confounders in our study.

Starting with 2,137 people, we first selected those of normal weight ($18.5 < BMI < 25$ kg/m²; N=1,259), of whom 800 were women, with ages ranging from 30 to 65 years. We selected a younger subset of 280 women aged between 30 and 40 years. Of these, 60 had predicted *AMY1* CN <=4 (LA) and 86 CN>=8 (HA) – which represent, respectively, the first and third quartile of the copy number distribution observed in this population. Availability of nine years of follow-up data in the D.E.S.I.R. cohort reduces the possibility that underlying (and undetected) predisposition to metabolic disorders was present in women belonging to the two selected groups, thus reducing potential confounding from factors that could have been disproportionately prevalent in one of the two genotype groups.

Fifty age-matched women were selected from each of the two LA/HA groups. No significant differences in self-reported food preference and intake nor sporting activity were detected between the two groups. Finally, serum salivary amylase levels were used to validate the genetic data and verify whether low and high copy number at *AMY1* was also reflected in high and low serum amylase levels.

MASS SPECTROMETRY

The metabolon analytical platform consisted of two separate ultra-high performance liquid chromatography/tandem mass spectrometry devices (UHPLC/MS/MS^2), optimized one for basic and acidic species, and one for GC/MS. Each sample was run simultaneously on the 3 analytic machines.

Metabolites were identified by automated comparison of the ion features in the experimental samples to a reference library of chemical standard entries that included retention time, molecular weight (m/z), preferred adducts, and in-source fragments as well as their associated MS/MS^2 spectra. This integrated platform enables the high-throughput collection and relative quantitative analysis of analytical data and identifies a large number and broad spectrum of molecules with a high degree of confidence. Analytical repeatability was assessed by calculating the median relative standard deviation (RSD) for the internal standards that were added to each sample prior to injection into the mass spectrometers. Values for instrument and process variability were below significance threshold (<5%).

Prior to statistical analysis, we applied a conservative missing value cut off of 20% to the raw area count values, which led to the exclusion of 40 metabolites. 407 biochemicals thus remained for further analysis. Missing values for the remaining metabolites were inputted using the observed minimum detection value. Normalization was performed by Median Fold Change Normalization and autoscaling applied prior to performing the data analysis.

TABLES

Table S1. Comparison of age, metabolic parameters, and serum amylase levels (total, pancreatic, salivary) between women included in the HA and LA groups. Values are reported as mean (SD). Potential differences in quantitative variables between groups were tested by non-parametric Wilcoxon tests. No significant difference in the variance of these variables (Levene test P value > 0.05) was observed between the HA and LA groups. The BMI distribution does not differ between the two groups (Wilcoxon rank sum test P value = 0.986), and adjusting the variables compared in table S1 for BMI had a negligible effect on the test results.

	High AMY1 CN	**High AMY1 CN**	**P value**
Age in years ± SD	36 ± 3.3	35 ± 3.2	ns
Average BMI (kg/m^2) ± SD	22 ± 2.0	21.5 ± 1.8	ns
Average HbA1c (%; mmol/mol) ± SD	5.00 ± 0.24 (31.26 ± 2.66)	4.93 ± 0.35 (31.22 ± 3.88)	ns
Fasting glucose (mg/dl)	4.89 ± 0.38	4.89 ± 0.37	ns
Serum total amylase (U/L) ± SD	77.22 ± 36.34	50.82 ± 17.7	6.40×10^{-6}
Serum pancreatic amylase (U/L) ± SD	32.8 ± 23.48	26.58 ± 10.56	0.025
Serum salivary amylase (U/L) ± SD	44.39 ±19.58	24.24 ± 15.04	6.04×10^{-8}

Table S2. Top 5 Molecular and cellular Functions categories identified by IPA knowledgebase analysis using the discriminant metabolites identified in this study. 32 out of 41 molecules were mapped into the IPA knowledgebase analysis. The top 5 Molecular and cellular Functions included (with $P<0.05$) Cell Death and survival, Carbohydrate metabolism, Energy production, Lipid metabolism, Small molecule Biochemistry. The top functional annotations shared by the latest 4 categories where the Oxidation of glucose-6-phosphate (P=3.12×10-8), Oxidation of lipid (P=1.32×10-6) and Accumulation of acylglycerol (P=6.19×10-8) and triacylglycerol (P=2.85×10-6).

Name	P value range	# Molecules
Cell Death and survival	4.48×10^{-2} - 2.23×10^{-9}	10
Carbohydrate metabolism	3.76×10^{-2} - 3.12×10^{-8}	7
Energy production	3.76×10^{-2} - 3.12×10^{-8}	3
Small molecule biochemistry	4.10×10^{-2} - 3.12×10^{-8}	12
Lipid metabolism	4.10×10^{-2} - 6.19×10^{-8}	9

FIGURES

Figure S1. Whisker diagram of salivary (**A**) and pancreatic (**B**) amylase levels measured in serum samples obtained from individuals displaying low (LA; n=50) or High (HA; n=50) copies of the AMY1 gene. The concentration of salivary amylase was obtained by subtracting the pancreatic amylase level from the total amylase level.